FG-PersonX: A Dataset for Fine-Grained Person Re-identification

Ruotian Zhang†, Yuchi Liu†, Xiaoxiao Sun†, Vishal Choudhary‡, Hock Beng Lim‡, Ngai-Man Cheung‡
†Australian National University
‡ Center for Smart Systems, Singapore University of Technology and Design
[Ruotian.Zhang, Yuchi.Liu, Xiaoxiao.Sun]@anu.edu.au, [vishal_choudhary, hockbeng_lim, ngaiman_cheung]@sutd.edu.sg

Abstract

In this work, we develop a new large-scale synthetic dataset, named "FG-PersonX", to support the study of fine-grained person re-identification (FG person re-ID). Existing person re-ID works usually focus on scenarios where the appearance of people, such as the color and style of their clothes, is distinguishable. However, person re-ID is more challenging for some scenarios, such as in schools, hospitals, and factories, where people wear similar clothes such as uniforms. To provide data for studying FG person re-ID, our dataset has three key aspects. First, it contains 1,000 identities of five groups: students, medical personnel (e.g., doctors and nurses), construction workers, sportsmen, and security personnel. Second, it has over 40K images generated under different indoor and outdoor scenes. Third, we provide a data synthesizing engine that can generate customized data in different environments by changing visual factors such as illumination, backgrounds, and viewpoints. We have performed extensive evaluations using FG-PersonX for benchmarking purposes. We believe this dataset will be beneficial for studies on both fine-grained recognition and person re-ID.

Figure 1: Re-ID results of IDE [21] on generic (**left**) and FG person re-ID datasets (**right**). The re-ID performance has a dramatic drop on the FG dataset compared with that on the generic dataset. The real-world data (**A**) are the Market-1501 and its subset. The synthetic data (**B**) are generated by using sampled models of our datasets.

1. Introduction

This work focuses on the fine-grained person re-identification (FG person re-ID) task. In an environment where people wear similar clothes such as uniforms, given a probe image (query), the goal is to find a gallery (database) of images that contains the same person.

Existing deep learning techniques have shown promising results in solving the general person re-identification problem [4, 14]. However, the performance of existing methods degrades rapidly when applied to scenarios where the variation in appearances is more nuanced and fine-grained [19]. Fig. 1 shows two validation experiments in which the performance of the same method will drop dramatically on fine-grained image data. The scenarios where people dress similarly in uniforms are commonly seen in schools, hospitals, factories, etc. Similar to fine-grained object recognition [17, 5, 12, 15, 7], *small inter-class* and *large intra-class variations* are the main challenges for FG person re-ID. However, the two tasks are different in several aspects. For example, the setting of the training and test sets is different. Specifically, the categories of the training set will not appear in the test set of FG person re-ID, while they are often the same in the training and test sets for FG object recognition. Therefore, FG person re-ID can be regarded as a special fine-grained recognition task.

Considering the potential application and research values of FG person re-ID, we build a large-scale synthetic dataset, FG-PersonX, for studying FG person re-ID. The reason for using synthetic data is that manually annotating people with similar appearances across different cameras is expensive and error-prone. For example, it is hard for humans to distinguish different nurses when they wear similar uniforms. Also, the appearance of nurses with the same identity may change dramatically across cameras (*e.g.*, variances of pose, illumination, and background), which will increase the difficulty of annotation, and even cause incorrect labels. Therefore, we use the approach of generating synthetic data.

Inspired by recent works using synthetic data [20, 16, 13, 18], we develop a system based on the Unity engine to generate images automatically for different identities with accurate annotations. The system includes 1000 3D person models of students, medical personnel, construction workers, sportsmen, and security personnel, as well as 5 scenes. The identity models of each category of characters wear uniforms, so their appearances are very similar. To meet different requirements, we provide a controllable interface that enables users to add new scenes, models and modify the parameters controlling visual factors, such as illumination and camera location. Based on this system, the FG-PersonX dataset is generated with over 300K images. We will discuss the details of the system and FG-PersonX dataset in Section 2.

For benchmarking purpose, we evaluate many representative person re-ID techniques on the FG-PersonX dataset. Experimental results show that our dataset is indicative of the performance of different techniques.

2. FG-PersonX

2.1. Data Synthesizing System

By using the Unity[1] [10] engine, we create a controllable system to generate images with 3D models. The assets include 3D models of person and scene. We also provide interfaces with editable parameters that can be used to modify commonly studied visual factors in person re-ID, such as viewpoint, illumination, and background

Identities: There are 1,000 hand-crafted identities of 5 groups including students, medical personnel (doctors and nurses), construction workers, sportsmen, and security personnel. All of them are designed to wear uniforms, and have only subtle differences to distinguish them. To ensure diversity, we hand-craft the person models with different genders, skin colors, ages, hairstyles, *etc.* Fig. 2 **A** shows sample image models of different identities.

[1]https://unity.com/

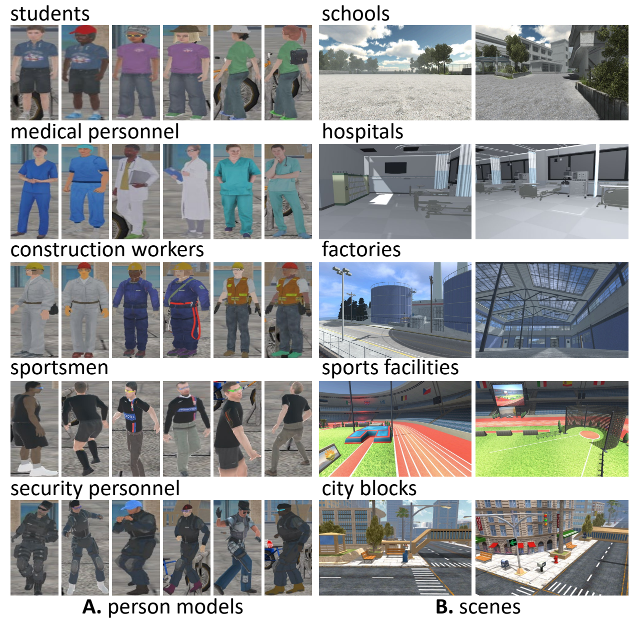

Figure 2: Examples of **A.** 3D person models and **B.** scenes used to generate FG-PersonX. These models and scenes can be used to generate diverse images for person re-ID.

Scenes: To simulate realistic environments, five scenes are provided corresponding to different groups, including schools, hospitals, factories, sports facilities and city blocks (shown in Fig. 2 **B**). Furthermore, we provide an interface that enables users to add other scenes into the system based on their requirements.

Visual Factors: The options for illumination include directional light (sunlight), point light, spotlight, and area light. More complex illuminations can be created by mixing different kinds of light and reconsidering object reflection. The backgrounds will change with the field of view (FoV) of the camera and the position of the person in the 3D scene. More variance in the 3D scenes means images can be generated with more diverse backgrounds. As mentioned above, we have provided five scenes for users to synthesize image data with various backgrounds. The viewpoint of a person can be changed from $0°$ to $360°$, and can be accurate to $1°$.

2.2. Image Dataset

Based on the above system, we have created the FG-PersonX image dataset for research purposes. The settings of the datasets are introduced below.

Cameras: Person re-ID is a cross-camera image retrieval task and the arrangement of cameras matters in a re-ID dataset. Here, We arrange 2 cameras at different po-

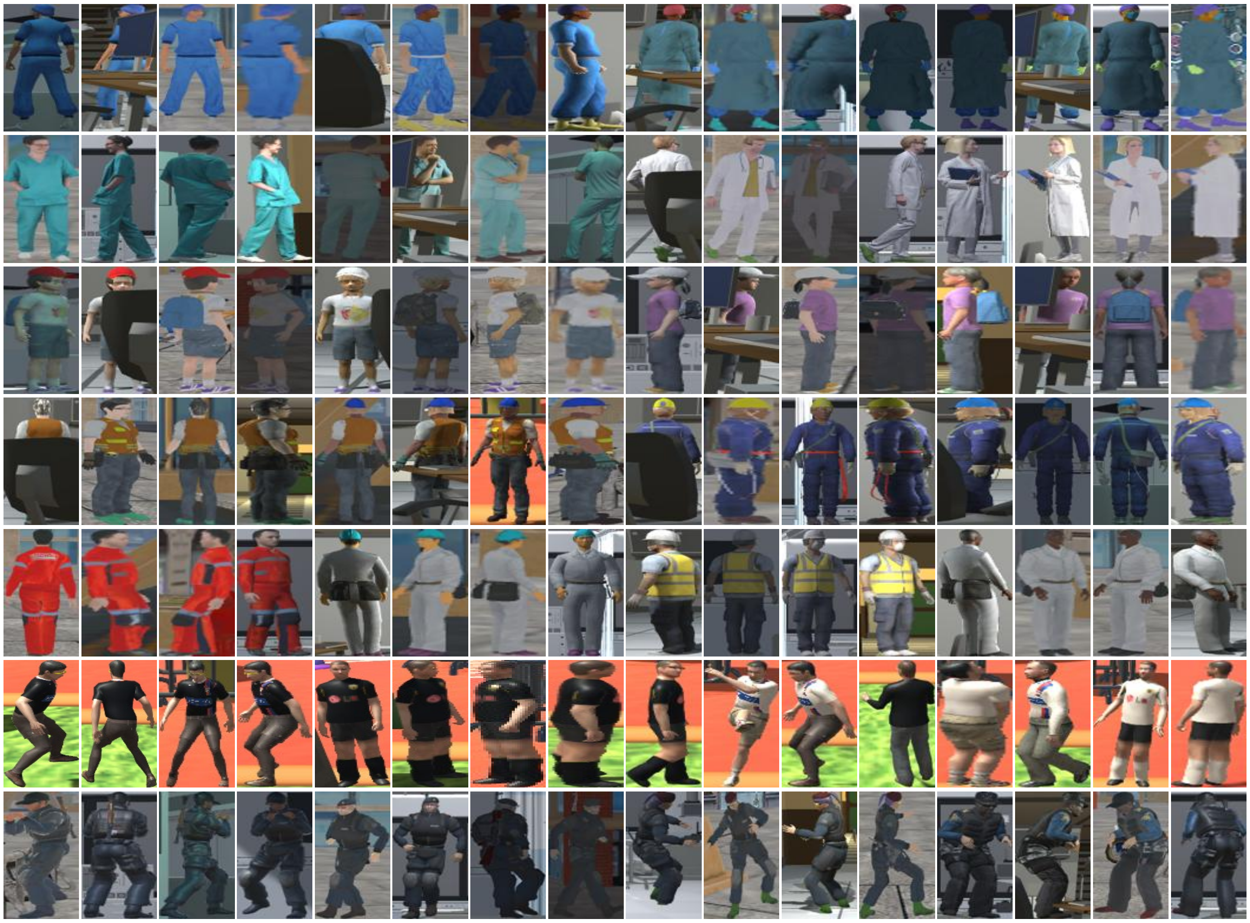

Figure 3: Sample images of the FG-PersonX dataset. They are generated with different backgrounds, illuminations and resolutions. Most of these models have different poses and some of images are generated with random occlusions.

	dataset	#identity	#box	#cam.	Released?
Generic	Market-1501 [22]	1,501	32,668	6	Y
	MARS [21]	1,261	1,191,003	6	Y
	CUHK03 [6]	1,467	14,096	2	Y
	Duke [11]	1,404	36,411	8	Y
	SOMAset [2]	50	100,000	250	N
	SyRI [1]	100	1,680,000	–	N
	PersonX [13]	1,266	arbitrary	arbitrary	Y
	UnrealPerson [20]	3,000	120,000	34	Y
	RandPerson [16]	8,000	228,655	19	Y
FG	FGPR [19]	358	134,696	5	N
	FG-PersonX-sys	1,000	arbitrary	arbitrary	TBD
	FG-PersonX-img	1,000	42,291	10	TBD

Table 1: Comparison of FG-PersonX, FGPR and other person re-ID datasets. "arbitrary" means users can generate any amount of data and add any number of cameras by modifying the parameters in the system. FG-PersonX-img is an image dataset we generated for experiments.

sitions in each scene for a total of 10 cameras. Users can modify the number of cameras arbitrarily to suit their requirements. The resolution of the cameras is set to 1024×768.

Viewpoints: For each identity, we change its rotation angle (relative to the camera) in $10°$ to capture 35 images for each camera, and then randomly sample 5 or 6 images to create our dataset. There are 49,291 images generated in the current FG-PersonX dataset. Sample images of the FG-PersonX dataset are shown in Fig. 3.

A comparison of FG-PersonX and an existing FG re-ID dataset, FGPR, as well as some generic person re-ID datasets is presented in Table 1. To our knowledge, FGRP [19] is the only existing FG re-ID dataset, but it has a limited number of IDs. Other synthetic datasets [20, 14, 16] are designed for generic situations, so they cannot be used for FG person re-ID. In comparison, FG-PersonX has more IDs, and users can generate customized data using our system. Furthermore, it can be edited/extended not only for this

Person_reID_baseline [23]	reid-strong-baseline [9]	deep-person-reid [25]
IDE, PCB, DenseNet	ResNet18, ResNet34 ResNet50, SeResNet50, SeResNeXt50, IBN-Net50	osnet-x0-25, osnet-x0-50, osnet-x0-75, osnet-x1-0, resnet50-fc512

Table 2: Names of methods evaluated on the FG-PersonX.

study but also for future research in this area. The release of FG-PersonX is to be determined (TBD).

3. Experiment

We evaluate commonly used techniques of person re-ID using the FG-PersonX dataset for benchmarking purposes. Data of different roles are mixed as the training and testing data.

3.1. Experiment Setting

Models: Table 2 shows the names of evaluated techniques/models in our experiment. They are selected from three highly-rated Github repositories: Person_reID_baseline[2], reid-strong-baseline[3] and deep-person-reid[4]. The hyperparameters of training models are set based on the original settings of these projects.

Evaluation Metrics: In our evaluation, we use the Cumulative matching characteristics (CMC) and mean average precision (mAP), which are standard metrics for person re-ID research [22, 8]. In CMC, rank-k accuracy indicates the probability that a queried identity appears in the top-k ranked candidates list, and ranks 1, 5 and 10 are commonly reported in the literature. Thus, we report the Rank1, Rank5, Rank10, and mAP of different techniques on the FG-PersonX dataset.

3.2. Evaluation on FG-PersonX

We compare different techniques using the FG-PersonX dataset, and the results are reported in Table 3. Based on the results, there are several observations. First, FG-PersonX is suitable for model evaluation because the comparison results on it are consistent with those on other datasets [13, 14]. For example, it has been reported in many papers [3, 24] that PCB performs better than IDE. On FG-PersonX, PCB gets a mAP score of 68.5% which is higher than 64.3% of IDE. Meanwhile, Resnet50 gets higher mAP and CMC scores than ResNet18 and Resnet34 on FG-PersonX, which is consistent with the results reported in other papers [9]. Second, the local feature helps to recognize the fine-grained person. PCB uses part-level features and outperforms the models learning using global features, such as IDE and DenseNet. Third, the result on FG-PersonX is hard to improve, because improvements of different methods are smaller on our dataset than on other datasets. For example, methods in Table 3 can get 65% to 85% mAP on the Market-1501 [9, 25], but the highest mAP on FG-PersonX is 73.5%.

Thus, the proposed FG-PersonX dataset can be used to study the performance of different techniques for FG person re-ID. Furthermore, more complicated images can be generated by using our system to create more challenging datasets.

Methods	Rank1	Rank5	Rank10	mAP
IDE	75.1	86.2	88.1	64.3
PCB	77.8	86.9	88.4	68.5
DenseNet	75.0	86.6	88.5	63.8
ResNet18	75.8	86.4	88.2	66.2
ResNet34	75.4	86.3	88.2	66.5
ResNet50	77.3	86.7	88.3	69.1
SeResNet50	76.9	86.6	88.5	69.6
SeResNetXt50	78.1	87.4	88.7	71.9
IBN-Net50	77.2	86.9	89.4	69.7
osnet-x0-25	79.1	87.3	88.8	70.2
osnet-x0-50	79.8	87.4	88.6	71.9
osnet-x0-75	80.5	87.5	88.9	73.5
osnet-x1-0	80.3	87.6	88.7	73.4
resnet50-fc512	79.7	87.4	88.5	71.2

Table 3: Evaluation results (Rank-K(%) and mAP(%)) of different techniques on the FG-PersonX dataset.

4. Conclusion

We have developed the FG-PersonX dataset which overcomes the burden of costly data annotation and provides data for FG person re-ID studies. Our data synthesizing system can generate image data under controllable cameras, backgrounds, illuminations, *etc*. For benchmarking purposes, several commonly used re-ID techniques were evaluated on FG-PersonX. Experimental results show that FG-PersonX is indicative of the performance of different techniques. Thus, future work on FG person re-ID can be conducted based on this dataset.

Acknowledgments

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-100E-2019-040) and ST Engineering Electronics Ltd.

[2] https://github.com/layumi/Person_reID_baseline_pytorch
[3] https://github.com/michuanhaohao/reid-strong-baseline
[4] https://github.com/KaiyangZhou/deep-person-reid

References

[1] Slawomir Bak, Peter Carr, and Jean-Francois Lalonde. Domain adaptation through synthesis for unsupervised person re-identification. In *ECCV*, 2018. 3

[2] Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. *Computer Vision and Image Understanding*, 167:50–62, 2018. 3

[3] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, 2018. 4

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[5] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Fei-Fei Li. Novel dataset for fine-grained image categorization: Stanford dogs. In *Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, 2011. 1

[6] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 3

[7] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear cnn models for fine-grained visual recognition. In *ICCV*, pages 1449–1457, 2015. 1

[8] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 1–6, 2016. 4

[9] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 4

[10] John Riccitiello. John riccitiello sets out to identify the engine of growth for unity technologies (interview). *VentureBeat. Interview with Dean Takahashi. Retrieved January*, 2015. 2

[11] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV*, pages 17–35, 2016. 3

[12] Shihan Su, Yuxin Chen, Oisin Mac Aodha, Pietro Perona, and Yisong Yue. Interpretable machine teaching via feature feedback. 2017. 1

[13] Xiaoxiao Sun and Liang Zheng. Dissecting person re-identification from the viewpoint of viewpoint. In *CVPR*, 2019. 2, 3, 4

[14] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, pages 480–496, 2018. 1, 3, 4

[15] Grant Van Horn, Steve Branson, Ryan Farrell, Scott Haber, Jessie Barry, Panos Ipeirotis, Pietro Perona, and Serge Belongie. Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In *CVPR*, pages 595–604, 2015. 1

[16] Yanan Wang, Shengcai Liao, and Ling Shao. Surpassing real-world source training data: Random 3d characters for generalizable person re-identification. In *ACMMM*, 2020. 2, 3

[17] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010. 1

[18] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. Simulating content consistent vehicle datasets with attribute descent. In *ECCV*, 2020. 2

[19] Jiahang Yin, Ancong Wu, and Wei-Shi Zheng. Fine-grained person re-identification. *International journal of computer vision*, pages 1–19, 2020. 1, 3

[20] Tianyu Zhang, Lingxi Xie, Longhui Wei, Zijie Zhuang, Yongfei Zhang, Bo Li, and Qi Tian. Unrealperson: An adaptive pipeline towards costless person re-identification. *arXiv preprint arXiv:2012.04268*, 2020. 2, 3

[21] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, pages 868–884, 2016. 1, 3

[22] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 3, 4

[23] Zhedong Zheng, Liang Zheng, and Yi Yang. A discriminatively learned cnn embedding for person reidentification. *TOMM*, 14(1):13, 2018. 4

[24] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 4

[25] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *ICCV*, 2019. 4